INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF RADIOLOGICAL PROTECTION

J. Radiol. Prot. **22** (2002) 141–148 PII: S0952-4746(02)35522-8

The potential for bias in Cohen's ecological analysis of lung cancer and residential radon

Jay H Lubin

Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, EPS/8042, 6120 Executive Blvd, Rockville, MD 20892-7244, USA

E-mail: lubinj@mail.nih.gov

Received 7 November 2001, in final form 28 January 2002, accepted for publication 5 March 2002
Published 30 May 2002
Online at stacks.iop.org/JRP/22/141

Abstract

Cohen's ecological analysis of US lung cancer mortality rates and mean county radon concentration shows decreasing mortality rates with increasing radon concentration (Cohen 1995 *Health Phys.* **68** 157–74). The results prompted his rejection of the linear-no-threshold (LNT) model for radon and lung cancer. Although several authors have demonstrated that risk patterns in ecological analyses provide no inferential value for assessment of risk to individuals, Cohen advances two arguments in a recent response to Darby and Doll (2000 *J. Radiol. Prot.* **20** 221–2) who suggest Cohen's results are and will always be burdened by the ecological fallacy. Cohen asserts that the ecological fallacy does not apply when testing the LNT model, for which average exposure determines average risk, and that the influence of confounding factors is obviated by the use of large numbers of stratification variables. These assertions are erroneous. Average dose determines average risk only for models which are linear in all covariates, in which case ecological analyses are valid. However, lung cancer risk and radon exposure, while linear in the relative risk, are not linearly related to the scale of absolute risk, and thus Cohen's rejection of the LNT model is based on a false premise of linearity. In addition, it is demonstrated that the deleterious association for radon and lung cancer observed in residential and miner studies is consistent with negative trends from ecological studies, of the type described by Cohen.

1. Introduction

Cohen [1] has published ecological analyses of US lung cancer mortality rates and estimates of mean county radon concentrations showing a general decrease in mortality rates with increasing radon levels, which prompted his rejection of the linear-no-threshold (LNT) model for radon and lung cancer. The irrelevance of ecological observations to risk of lung cancer to individuals residentially exposed to radon has been the subject of several exchanges [2–12]. In a response

to a recent comparative analysis of ecological and analytical case-control data by Darby and Doll [13], Cohen [14] advances two arguments in his rejection of the LNT model. While accepting that '*average* exposure does not, in general, determine the *average* risk' (Cohen's emphasis), Cohen asserts that the ecological fallacy does not apply since he is only testing the LNT model, for which average exposure does determine average risk. Cohen further asserts that the influence of confounding factors (CF) on the lung cancer and radon association is necessarily obviated by his use of large numbers of stratification (i.e., regressor) variables. Both of these assertions are erroneous.

2. Linearity of the risk function for lung cancer and radon

In this section, we show that Cohen's ecological model is not the natural extension to the county for the risk of lung cancer to individuals. This misspecification in the aggregate model is critical since it intrinsically invalidates the conceptual basis of Cohen's 'test of the LNT hypothesis'. A statistical test of a hypothesis is a probability statement about the plausibility of a null value for a parameter in comparison to the defined alternative. For example, in a model $f(x, y; \alpha, \beta)$ with factors x and y, a test of the null hypothesis $\beta = \beta_0$ compares the plausibility of $f(x, y; \hat{\alpha}_0, \beta_0)$ in comparison to $f(x, y; \hat{\alpha}_1, \beta_1)$ where $\beta = \beta_1$ is the value under the alternative and $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are estimates of α under null and alternative hypotheses, respectively. However, this evaluation proceeds under a specific model and is informative only to the extent that the model under the alternative represents the true state of nature, i.e., the true relationship. In Cohen's case, he assumes a linear relationship for the county lung cancer mortality rate and the mean county radon concentration, and defines the null hypothesis as a positive slope parameter compatible with risk estimates from miner analyses. Cohen's analysis shows—quite convincingly—that age-adjusted county lung cancer mortality rates are inversely related to his estimated county radon concentrations and therefore incompatible with this null. On that basis, he concludes that the null should be rejected. However, his rejection of the null is applicable only within the context of the model used in the ecological regression. Since we show below that the form of his model is not compatible with current understanding, there is little scientific information that derives from this rejection of the null. As an extreme example, suppose one were to assume that county lung cancer rates are sinusoidal with radon concentration. An analysis of data that rejects the associated parameters provides no insight for individual risk, only indicating that county rates are not compatible with a sine pattern.

Thus, Cohen's first assertion regarding 'testing' the linearity of the exposure response depends on the compatibility of the ecological model with the individual-level relationship. Inference based on aggregate data is valid if the individual-level model is linear in the absolute risk in all risk factors with no interactions. This is not the case with radon, where analysis of miners indicates linearity with cumulative exposure, but linearity in the relative risk (RR) [15]. On an absolute risk scale, the radon effect varies with age at risk, time since exposure, exposure rate, smoking status, and sex. Since neither the individual nor the county-level models for absolute risk are linear in radon exposure, Cohen's modelling does not eliminate the potential for ecological bias.

Analyses have shown that the lung cancer risk function for an individual is not linear in radon exposure and other risk factors [15]. Analyses of miner data demonstrate linearity in the RR with radon exposure, with a slope parameter dependent on age, cigarette smoking status, and other factors. Consequently, the mean lung cancer rate in a group of heterogeneous (in age, smoking status, and other factors) individuals is not linear in mean radon exposure. Thus, lung cancer rates cannot be compared unless the joint distributions of radon, age, smoking, and other risk variables within groups are the same or are known.

To demonstrate, assume the radon exposure and cigarette use are the only risk factors. A reasonable first-order approximation of lung cancer risk, r, for individual j in county i is

$$r(w_{ij}, s_{ij}) = \gamma_0 \theta^{s_{ij}} (1 + \beta w_{ij}) \tag{1}$$

where w_{ij} is radon concentration assumed to act over a fixed time period, and s_{ij} is the smoking status (with $s = 1$ denoting smoker and $s = 0$ denoting non-smoker). The unknown parameters γ_0, θ, and β represent the background lung cancer risk in non-exposed non-smokers, the RR of smoking, and the true (linear) excess RR per unit radon concentration, respectively. If equation (1) is aggregated over individuals in a county, then the induced regression model for the mean lung cancer rate in the ith county is

$$\bar{r}_i = \gamma_0 (P_{0i} + \theta P_{1i})(1 + \beta \tilde{w}_i) \tag{2}$$

where P_{1i} $(=1 - P_{0i})$ is the proportion of smokers and $\tilde{w}_i$ is the 'risk-weighted' mean radon concentration, given by

$$\tilde{w}_i = \frac{P_{0i}}{P_{0i} + \theta P_{1i}} \bar{w}_{0i} + \frac{\theta P_{1i}}{P_{0i} + \theta P_{1i}} \bar{w}_{1i}. \tag{3}$$

It is important to note that $\bar{w}_{0i}$ and $\bar{w}_{1i}$ are mean radon concentrations within the county for non-smokers and smokers, respectively, information not available for ecological regression. Equation (2) represents the true association at the county level, and so the true β can be estimated if the $\tilde{w}_i$ are known for all counties. In contrast to the 'risk-weighted' $\tilde{w}_i$, the simple county mean radon used in ecological regression is $\bar{w}_i = P_0 \bar{w}_{0i} + P_1 \bar{w}_{1i}$. Adding and subtracting $\beta \bar{w}_i$, equation (2) can be rewritten as

$$\bar{r}_i = \gamma_0 (P_{0i} + \theta P_{1i}) \left\{ 1 + \beta \bar{w}_i + \beta \frac{\theta - 1}{P_{0i} + \theta P_{1i}} \operatorname{cov}(w_i, s_i) \right\} \tag{4}$$

where $\operatorname{cov}(w_i, s_i) = P_{0i} P_{1i} (\bar{w}_{1i} - \bar{w}_{0i})$ is the covariance of radon and smoking status within the ith county. If $\theta = 1$ or $\operatorname{cov}(w_i, s_i) = 0$ for all counties, then the county lung cancer rate $\bar{r}_i$ is linear in the RR with $\bar{w}_i$, and β can be estimated. If these conditions are violated, then neither the risk model for an individual, equation (1), nor the induced county-level model, equation (4), is linear in the risk factors.

Equations (1)–(4) demonstrate that, even with a single factor such as smoking and ignoring the numerous other risk factors, linearity in the RR for radon for individuals does not translate into linearity of the absolute risk or linearity of the RR at the county level. Cohen's first argument is therefore based on an erroneous premise, namely linearity of risk factors, and therefore his 'rejection' of the LNT model is noninformative.

Equation (4) offers guidance on how one might carry out an unbiased ecological regression [16]. Consider the major risk factors for lung cancer: smoking (duration, cigarettes per day, time since cessation, filtered or non-filtered type of cigarette), age, sex, various occupational exposures, air pollution, and nutritional status [17]. Many of these are probably correlated directly or indirectly with radon concentration. One approach estimates the joint covariances of radon and these risk factors using random population samples within each county [18, 19]. A second approach aggregates units such that radon concentration and all risk factors are uncorrelated within each unit. An unbiased ecological regression would therefore require data summarised at a very detailed level—that is, the lung cancer mortality rate and radon concentration within units defined by the cross-tabulation of sex, categories of age, smoking characteristics, etc. This detail is not typically available, and cannot be approximated by adding large numbers of regression covariates, as is attempted by Cohen.

3. The potential impact of cross-level bias

Cohen's second assertion confuses confounding at the aggregate level and at the individual level. The effect of confounding at the individual level is bounded by the magnitude of the association of the CF with disease. For example, a twofold CF induces a twofold distortion of the risk factor if and only if the CF is perfectly correlated with the risk factor; otherwise the impact of the CF on a risk factor must be less. In an ecological analysis, factors can confound at the aggregate level or at the individual level. County-level confounders are controlled by the addition of covariates in the regression model. Cohen's adjustment for county-level confounding using county-level regressor variables is extensive. However, in spite of Cohen's extreme adjustment, one cannot control for individual-level confounding by adding county-level variables. Moreover, cross-level bias is potentially unbounded and small within-county correlations can markedly effect the county-level relationship [8, 9].

The difference between the simple county mean $\bar{w}_i$ and $\tilde{w}_i$ is the source of the ecological bias, and derives from the variation of $\mathrm{cov}(w_i, s_i)$ across counties. One can re-express the issue and ask whether cross-level bias can explain Cohen's result of a statistically significant protective effect of radon and yet still be consistent with a deleterious effect of radon. This was demonstrated by Lubin [8] for two counties, and below it is demonstrated for all counties. Suppose radon risk is homogeneous across ages and depends only on radon level and smoking status, as in equation (1). Using Cohen's data on age-adjusted county lung cancer mortality rates for males ($n = 1599$ counties with non-zero rates), proportions of smokers, and mean radon concentrations (in Bq m^{-3}), $\bar{w}_i$, the fitted equation is

$$\hat{Y}_i = \hat{\gamma}_{0e}(P_{0i} + \hat{\theta}_e P_{1i})[1 + \hat{\beta}_{1e}\bar{w}_i + \hat{\beta}_{2e}(\bar{w}_i)^2] \tag{5}$$

where $\bar{w}_i$ is in units of 100 Bq m^{-3}, $\hat{Y}_i$ is the predicted value, $\hat{\gamma}_{0e} = 0.000\,15$, $\hat{\theta}_e = 8.828$, $\hat{\beta}_{1e} = -0.3794$, and $\hat{\beta}_{2e} = 0.1256$. Cohen includes the statistically significant quadratic term in $\bar{w}_i$ and we do as well.

Now, suppose the fitted $\hat{Y}_i$ are the county lung cancer rates, instead of the actual $\bar{r}_i$, and assume that equation (5) represents the true relationship perfectly. The RR for radon is $\hat{Y}_i/[0.000\,15(P_{0i} + 8.28P_{1i})] = [1 - 0.3794\bar{w}_i + 0.1256(\bar{w}_i)^2]$ (figure 1, panel A). Since the fit of the model is perfect by construction, there is no residual variation and therefore no additional county-level regression variables can improve the fit.

A judicious choice of $\bar{w}_{0i}$, where $\bar{w}_{1i}$ is defined by the constraint $\bar{w}_i = P_0\bar{w}_{0i} + P_1\bar{w}_{1i}$, preserves both the relationship in equation (5) and results in parameter estimates for equation (2) which are consistent with analytic studies. For each i, we equate equations (2) and (5), and solve for $\bar{w}_{0i}$ to obtain

$$\bar{w}_{0i} = \frac{1}{P_{0i}}\left[\frac{B_{ei} - B_{ti} - \{B_{ti}\beta(\frac{\theta}{P_{0i}+\theta P_{1i}}) - B_{ei}(\hat{\beta}_{1e} + \hat{\beta}_{2e}\bar{w}_i)\}\bar{w}_i}{-B_{ti}\beta\frac{(\theta-1)}{P_{0i}+\theta P_{1i}}}\right] \tag{6}$$

where $B_{ei} = \hat{\gamma}_{0e}(P_{0i} + \hat{\theta}_e P_{1i})$ and $B_{ti} = \gamma_0(P_{0i} + \theta P_{1i})$ are the estimated (based on the ecological regression) and true background rates, respectively. In this expression, $\bar{w}_{0i}$, the mean radon concentration among people who have never smoked for county i, is a function of the county-level data, P_{0i}, P_{1i}, and $\bar{w}_i$, the ecological regression estimates $\hat{\gamma}_{0e}$, $\hat{\theta}_e$, $\hat{\beta}_{1e}$, and $\hat{\beta}_{2e}$, and the true but unknown parameters from equation (2), γ_0, θ, and β. The goal is to specify values for γ_0, θ, and β, which are consistent with analytic studies of residential radon and lung cancer and result in values for $\bar{w}_{0i}$ and $\bar{w}_{1i}$ that satisfy equation (5). Cohen's ecological regression results are thus shown to be compatible with analytic studies, and highlights that ecological analyses provide little information on individual risk from radon exposure.

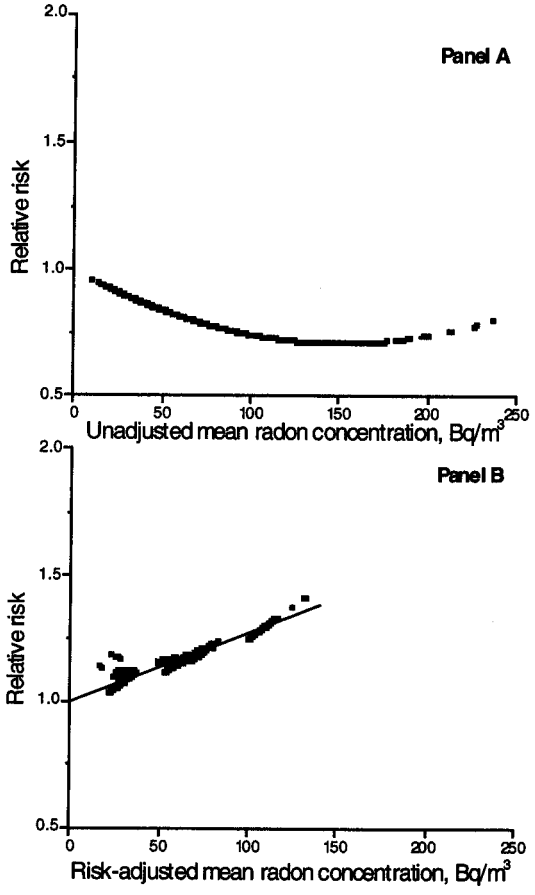

Figure 1. Using fitted lung cancer mortality rates ($\hat{Y}_i$), the RR for mean county radon concentration ($\bar{w}_i$) based on the ecological regression model, equation (5) (panel A), and for 'risk-adjusted' county radon concentration ($\tilde{w}_i$) based on the induced county-level true risk model, equation (2) (panel B).

First note that in equation (5), the risk for county i with P_{1i} smokers relative to county j with P_{1j} smokers, given that the counties have equal radon levels, $\bar{w}_i = \bar{w}_j$, is $(P_{0i} + \hat{\theta}_e P_{1i})/(P_{0j} + \hat{\theta}_e P_{1j})$, and independent of radon level. In contrast, for the true county-level model, the relative smoking effect is

$$\frac{(P_{0i} + \theta P_{1i})}{(P_{0j} + \theta P_{1j})} \frac{1 + \frac{\beta(\theta-1)\,\mathrm{cov}(w_i, s_i)}{(P_{0i}+\theta P_{1i})(1+\beta\bar{w}_i)}}{1 + \frac{\beta(\theta-1)\,\mathrm{cov}(w_j, s_j)}{(P_{0j}+\theta P_{1j})(1+\beta\bar{w}_j)}}$$

which for $\bar{w}_i = \bar{w}_j$ varies with $\mathrm{cov}(w, s)$. We took this into account when specifying the parameters in expression (6) used for creating $\tilde{w}_{0i}$.

For expression (6), we set $\gamma_0 = 0.000\,066$, $\beta = 0.68/100$ Bq m^{-3}, and $\theta = 15$, 12, 11, 10, and 8 for each quintile of radon concentration. For 120 counties (7.5%), equation (6) resulted in a non-positive value for $\tilde{w}_{0i}$. Those values were reset to the value for the county with the next lower radon concentration. Parameter estimates from fitting equation (2) are $\hat{\gamma}_0 = 0.000\,09$, $\hat{\theta} = 12.0$, 10.6, 9.8, 9.1, and 7.8 for each quintile, and $\hat{\beta} = 0.27$. The RR relationship for lung cancer and the 'risk-adjusted' radon concentration is shown in figure 1, panel B. It is seen that the 'risk-adjusted' county-level exposure-response relationship is linear

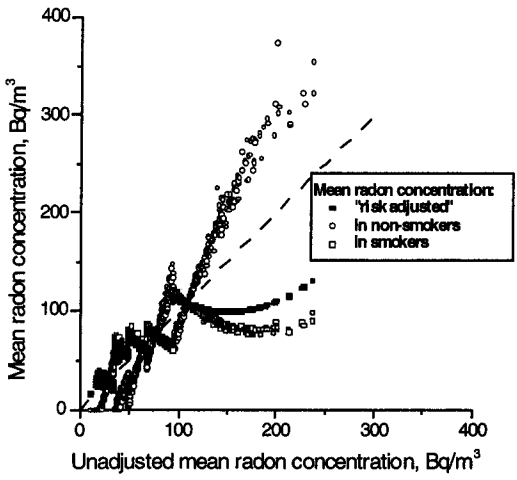

Figure 2. The relationship between the constructed 'risk-adjusted' mean radon concentration ($\tilde{w}_i$), mean radon concentration in non-smokers ($\bar{w}_{0i}$), mean radon concentration in smokers ($\bar{w}_{1i}$), and the unadjusted mean radon concentration. The dashed line is a 45° line representing no modification to the mean radon concentration.

and increasing. Moreover, the slope is compatible with previous estimates. Calculations based on miners who have had low exposure [20], residing in a typical residence for 35 years, result in an excess RR at 100 Bq m^{-3} of 0.41, similar to the estimate of 0.32 from a meta-analysis of residential studies [21] and to the estimate of 0.27 derived from the 'risk-adjusted' ecological regression.

The 'risk-adjusted' radon concentration for each county, $\tilde{w}_i$, is a weighted average of $\bar{w}_{0i}$ and $\bar{w}_{1i}$, the radon concentrations in non-smokers and in smokers, respectively. Figure 2 shows the relationship between $\bar{w}_{0i}$, $\bar{w}_{1i}$, $\tilde{w}_i$, and $\bar{r}_i$ where each discrete jump represents a different θ-value. The 45° broken line represents the mean of $\bar{w}_{0i}$ and $\bar{w}_{1i}$ weighted by P_{0i} and P_{1i}.

The construction of $\tilde{w}_i$ from $\bar{w}_i$ results in a shift towards the median, particularly in the upper extreme for the risk-adjusted mean. The lower quartile, median, and upper quartile for $\bar{w}_i$ are 36.6, 55.7, and 86.3 Bq m^{-3}, and for $\tilde{w}_i$ they are 55.9, 67.2, and 77.2 Bq m^{-3}. Assume that radon is log-normally distributed within county in smokers and in non-smokers with coefficient of variation 1.6, the value for all US houses [22]. For the constructed quantities, $\bar{w}_{0i}$ and $\bar{w}_{1i}$, the within-county correlation coefficients range from −0.33 to 0.41, with 1 (0.1%), 586 (36.7%), 759 (47.5%), and 253 (15.8%) correlation coefficients in the intervals [−0.5, −0.3), [−0.3, 0.0), [0.0, 0.3), and [0.3, 0.5), respectively.

4. Summary

In a simplified example, we fully account for the negative association in the county-level ecological regression while maintaining an appropriate risk model for individuals. Consequently, results from ecological studies may be interpreted as follows. Suppose counties have the same mean radon concentration and are identical for all county-level risk factors (i.e., the same mean ages, proportion of smokers, socioeconomic factors, etc). Even though all risk factors are identical at the county level, there is no logical inference that can be made about lung cancer rates for the counties. Differences in correlations among risk factors within counties can generate markedly different cancer rate patterns at the aggregate level. Conversely, no valid

inference for the radon and lung cancer association for individuals is possible on the basis of the relationship of county lung cancer rates. Thus, Cohen's results provide no information on risk to individuals, and do not necessarily contradict the 20 or more analytic studies which support an increasing association between radon and lung cancer.

Résumé

L'analyse écologique de Cohen sur les taux de mortalité aux Etats Unis par cancer du poumon et la concentration moyenne du radon dans le territoire, montre que les taux de mortalité diminuent quand la concentration du radon augmente (Cohen 1995 *Health Phys.* **68** 157–74). Le résultat le pousse à rejeter le modèle linéaire-sans-seuil (LSS) pour le radon et le cancer du poumon. Plusieurs auteurs ont démontré que les modèles de risque en analyses écologiques ne permettent pas de déduire une évaluation de la valeur du risque, pour un individu. Cependant, Cohen avance deux argumens dans une réponse récente à Darby et Doll (2000 *J. Radiol. Prot.* **20** 221–2) qui suggèrent que les résultats de Cohen sont et seront toujours hypothéqués par le sophisme écologique. Cohen prétend que le sophisme écologique ne s'applique pas quand on met à l'épreuve le modèle linéaire-sans-seuil, pour lequel l'exposition moyenne détermine le risque moyen, et que l'influence de facteurs de confusion est éliminée par l'emploi de grands nombres de variables de stratification. Ces assertions sont fausses. La dose moyenne ne détermine le risque moyen que dans le cas de modèles qui sont linéaires dans toutes leurs covariables; dans un tel cas, les analyses écologiques sont valables. Par contre, le risque de cancer du poumon et l'exposition au radon, bien que linéaires en ce qui concerne le risque relatif, ne sont pas liés de manière linéaire à l'échelle de risque absolu; le rejet par Cohen du modèle LSS est donc fondé sur une tautologie. De plus, il est démontré que l'association nuisible, en ce qui concerne le radon et le cancer du poumon, observée dans les études sur des batiments et sur des mines, est compatible avec les tendances négatives des études écologiques du type de celles décrites par Cohen.

Zusammenfassung

Cohens ökologische Analyse der Sterblichkeitsraten durch Lungenkrebs und der mittleren Radonkonzentration in den USA kommt zu dem Ergebnis, dass die Sterblichkeitsraten bei steigender Radonkonzentration fallen (Cohen 1995 *Health Phys.* **68** 157–74). Die Ergebnisse führten zu seiner Ablehnung des 'linear-no-threshold' (LNT) Modells für Radon und Lungenkrebs. Obwohl verschiedene Autoren nachgewiesen haben, dass Risikomuster in ökologischen Analysen keinen Inferenzwert für die Bewertung von Risiken für Menschen liefern, führt Cohen in einer kürzlich veröffentlichten Antwort auf Darby und Doll zwei Argumente an (2000 *J. Radiol. Prot.* **20** 221–2); Darby und Doll haben vorgeschlagen, dass Cohens Ergebnisse durch den ökologischen Trugschluss belastet sind und immer sein werden. Cohen behauptet, dass der ökologische Trugschluss auf Tests des LNT Modells nicht zutrifft, bei dem die durchschnittliche Strahlenbelastung das durchschnittliche Risiko bestimmt, und dass der Einfluss verwirrender Faktoren durch die Verwendung einer großen Anzahl von Schichtungsvariablen beseitigt wird. Diese Behauptungen sind falsch. Die durchschnittliche Dosis bestimmt das durchschnittliche Risiko nur bei Modellen, die in allen Kovarianten linear sind, so dass in diesem Fall die ökologischen Analysen gültig sind. Zwar sind Lungenkrebsrisiko und Radon-Strahlenbelastung beim relativen Risiko linear, sie stehen jedoch auf der Skala des absoluten Risikos nicht in einer linearen Beziehung und daher basiert Cohens Ablehnung des LNT-Modells auf einer Tautologie. Darüber hinaus wird gezeigt, dass

der schädliche Zusammenhang von Radon und Lungenkrebs, der in Studien über Wohnungen, sowie Bergleuten beobachtet wurde, vereinbar ist mit negativen Trends aus ökologischen Studien, von der Art, wie sie von Cohen beschrieben werden.

References

[1] Cohen B L 1995 Test of the linear–no threshold theory of radiation carcinogenesis for inhaled radon decay products *Health Phys.* **68** 157–74

[2] Piantadosi S, Byar D P and Green S B 1988 The ecologic fallacy *Am. J. Epidemiol.* **127** 893–904

[3] Greenland S and Robins J 1994 Invited commentary: ecologic studies—biases, misconceptions, and counterexamples *Am. J. Epidemiol.* **139** 747–60

[4] Piantadosi S 1994 Ecologic biases *Am. J. Epidemiol.* **139** 761–4

[5] Cohen B L 1998 Response to criticisms of Smith *et al Health Phys.* **75** 23–8

[6] Cohen B L 1998 Response to Lubin's proposed explanations of our discrepancy *Health Phys.* **75** 18–22

[7] Smith B J, Field R W and Lynch C F 1998 Residential ^{222}Rn exposure and lung cancer: testing the linear no-threshold theory with ecologic data *Health Phys.* **75** 11–17

[8] Lubin J H 1998 On the discrepancy between epidemiologic studies in individuals of lung cancer and residential radon and Cohen's ecologic regression *Health Phys.* **75** 4–10

[9] Lubin J H 1998 Rejoinder: Cohen's response to 'On the discrepancy between epidemiologic studies in individuals of lung cancer and residential radon and Cohen's ecologic regression' *Health Phys.* **75** 29–30

[10] Field R W, Smith B J and Lynch C F 1999 Comment on 'On the discrepancies between epidemiologic studies of lung cancer and residential radon and Cohen's ecologic regression' *Health Phys.* **76** 318–19

[11] Hei T K, Wu L J, Liu S X, Vannais D, Waldren C A and Randers-Pehrson G 1997 Mutagenic effects of a single and an exact number of alpha particles in mammalian cells *Proc. Natl Acad. Sci. USA* **94** 3765–70

[12] Lagarde F and Pershagen G 1999 Parallel analyses of individual and ecologic data on residential radon, cofactors, and lung cancer in Sweden *Am. J. Epidemiol.* **149** 268–74

[13] Darby S and Doll R 2000 Reply to 'Explaining the lung cancer versus radon exposure data for USA counties' *J. Radiol. Prot.* **20** 221–2

[14] Cohen B L 2001 Radon exposure and the risk of lung cancer *J. Radiol. Prot.* **21** 64–6

[15] National Research Council 1999 *Health Effects of Exposure to Radon (BEIR VI)* (Washington, DC: National Academy Press)

[16] Guthrie K A and Sheppard L 2001 Overcoming biases and misconceptions in ecological studies *J. R. Stat. Soc.* A **164** 141–54

[17] Blot W J and Fraumeni J F Jr 1996 Cancers of the lung and pleura *Cancer Epidemiology and Prevention* 2nd edn, ed D Schottenfeld and J F Fraumeni Jr (New York: Oxford University Press) pp 637–65

[18] Prentice R L and Sheppard L 1995 Aggregate data studies of disease risk factors *Biometrika* **82** 113–25

[19] Sheppard L and Prentice R L 1995 On the reliability and precision within and between population estimates of relative rate parameters *Biometrics* **51** 853–63

[20] Lubin J H *et al* 1997 Estimating lung cancer mortality from residential radon using data for low exposures of miners *Radiat. Res.* **147** 126–34

[21] Lubin J H 1999 Indoor radon and risk of lung cancer *(Proc. American Statistical Association Conf. on Radiation and Health (San Diego, CA, June 1998)); Radiat. Res.* **151** 105–7

[22] US Environmental Protection Agency 1992 *National Residential Radon Survey: Summary Report* EPA 402-R-92-011